


Nama: Ibnu Fajar Setiawan NIM: 065002000006	 Praktikum Data Warehouse	MODUL 12 Nama Dosen: Ir. Teddy Siswanto, MMSi
Hari/Tanggal: Senin, 20/06/2022		Nama Asisten Labratorium: 1. Azhar Rizki Zulma 065001900001 2. Nadiya Amanda Rizkania 064001900003

Proyek Akhir Praktikum IV

1. Teori Singkat

Data warehouse adalah jenis sistem manajemen data yang dirancang untuk memungkinkan dan mendukung kegiatan business intelligence (BI), terutama analitik. Gudang data semata-mata dimaksudkan untuk melakukan kueri dan analisis dan sering berisi sejumlah besar data historis. Data dalam gudang data biasanya berasal dari berbagai sumber seperti file log aplikasi dan aplikasi transaksi. Gudang data memusatkan dan mengkonsolidasikan sejumlah besar data dari berbagai sumber. Kemampuan analitisnya memungkinkan organisasi untuk memperoleh wawasan bisnis yang berharga dari data mereka untuk meningkatkan pengambilan keputusan. Seiring waktu, ia membangun catatan sejarah yang dapat sangat berharga bagi para ilmuwan data dan analisis bisnis. Karena kemampuan ini, gudang data dapat dianggap sebagai "sumber kebenaran tunggal" organisasi.

ETL Gabungan

ETL adalah singkatan dari Extract, Transform, Load. Sesuai namanya, Extract, Transform, Load merupakan proses yang meliputi ekstrak data dari beberapa sistem sumber, yang berlanjut kepada transformasi data (proses penghitungan, pembulatan, dan sebagainya) hingga akhirnya data ditampung (load) ke dalam sistem data warehouse. Data-data yang semula bercampur diolah menjadi data tunggal yang konsisten sehingga mudah untuk dilakukan analisis ketika dibutuhkan nantinya.

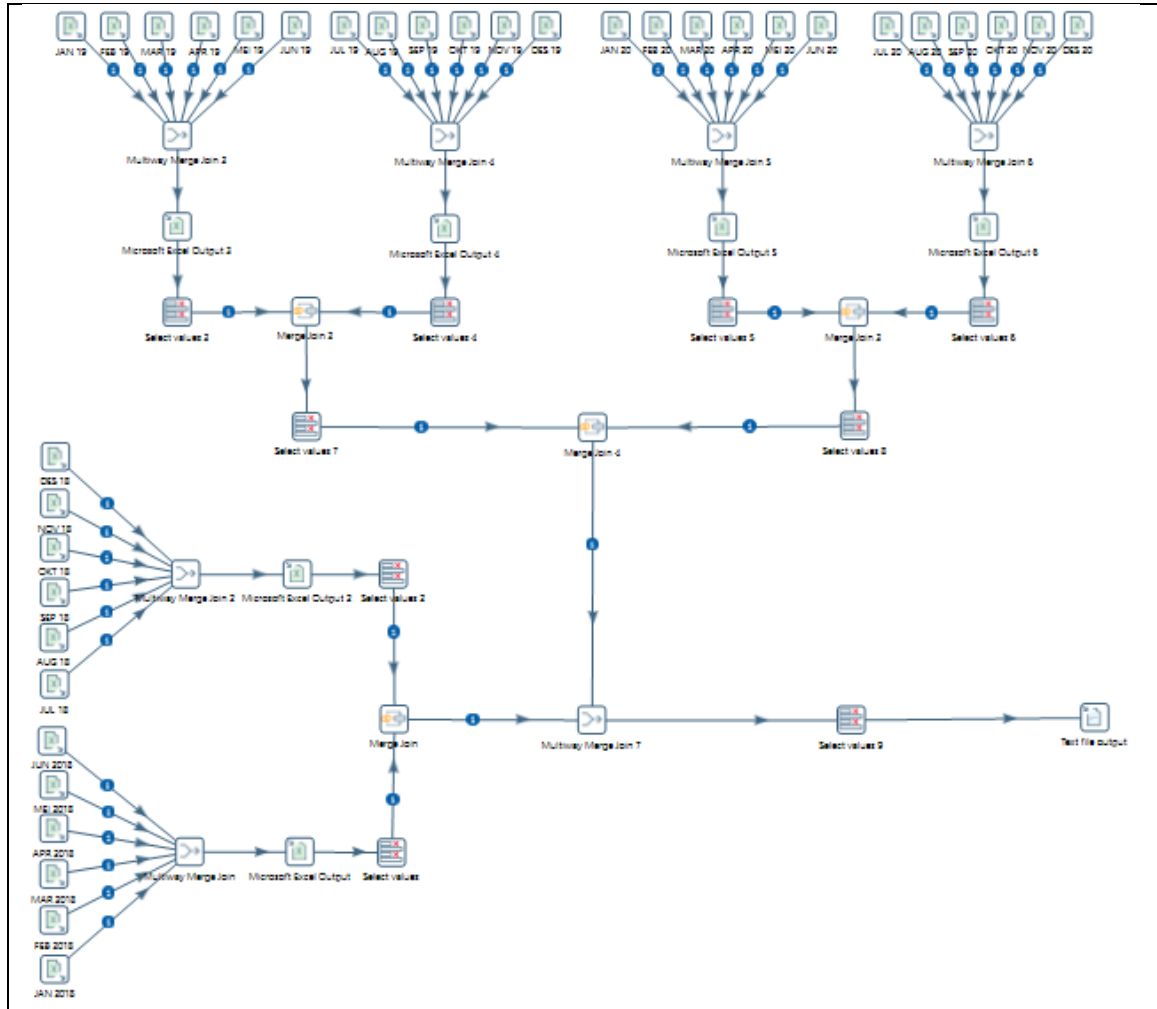
ETL menjadi pondasi untuk data analytics dan machine learning. Melalui sejumlah pengaturan, ETL dapat membersihkan dan mengorganisir data sesuai dengan yang diinginkan. Data laporan bulanan pun akan dapat dengan mudah dianalisis.



Berikut ini adalah langkah dalam melakukan ETL Gabungan menggunakan Spoon Pentaho:

1. Struktur

Buka Pentaho Spoon Data Integration lalu buat struktur seperti pada gambar dibawah ini lalu simpan dengan format nama ***Komoditi Cabai Merah Keriting.ktr***



2. Excel Input

Berikut ini adalah konfigurasi pada step *Microsoft Excel Input*, lakukan pada Jan 2018 – Des 2020 (Files, Sheets, Fields)

The image displays three sequential screenshots of the 'Microsoft Excel input' configuration window, showing the 'Files', 'Sheets', and 'Fields' tabs respectively.

Files Tab: The 'Step name' is 'JAN 2018'. The 'Spread sheet type (engine)' is 'Excel 97-2003 XLS (JXL)'. The 'File or directory' field is empty. The 'Regular Expression' and 'Exclude Regular Expression' fields are empty. The 'Password' field is empty. The 'Selected files' list shows one entry: '1 C:\Users\Azhar Rizky Zulma\OneDrive\Documents\Job\DWH\Praktikum Video Cleanshing and Filter D...'. The 'Accept filenames from previous steps' checkbox is unchecked. The 'Step to read filenames from' dropdown is set to 'Step to read filenames from'. The 'Field in the input to use as filename' dropdown is empty. The 'Show filename(s)...' button is visible.

Sheets Tab: The 'Step name' is 'JAN 2018'. The 'List of sheets to read' table is empty. The 'Get sheetname(s)...' button is visible.

Fields Tab: The 'Step name' is 'JAN 2018'. The 'Fields' table is populated with data. The 'Get fields from header row...' button is visible.

#	Name	Type	Length	Precision	Trim type	Repeat	Format	Currency	Decimal	Grouping
1	Pasar	String	-1	-1	none	N				
2	1	Number	-1	-1	none	N				
3	2	Number	-1	-1	none	N				
4	3	Number	-1	-1	none	N				
5	4	Number	-1	-1	none	N				
6	5	Number	-1	-1	none	N				
7	6	Number	-1	-1	none	N				
8	7	Number	-1	-1	none	N				
9	8	Number	-1	-1	none	N				
10	9	Number	-1	-1	none	N				
11	10	Number	-1	-1	none	N				
12	11	Number	-1	-1	none	N				
13	12	Number	-1	-1	none	N				
14	13	Number	-1	-1	none	N				



3. Multiway Merge Join

- Step selanjutnya adalah melakukan penggabungan data untuk menggabungkan data agar lebih mudah dalam mengolahnya. Step yang digunakan ialah Multiway Merge Join berikut ini konfigurasinya, lakukan pada Multiway Merge Join – Multiway Merge Join 6

The image displays two screenshots of the 'Multiway merge join' configuration window. The top window is titled 'Multiway merge join' and shows a list of input steps from JAN 2018 to JUN 2018, each with a 'Join Keys' field set to 'Pasar' and a 'Select Keys' button. The bottom window is titled 'Multiway merge join 2' and shows input steps from JUL 18 to DES 18, also with 'Join Keys' set to 'Pasar'. Both windows include a 'Join Keys' panel on the right with a table showing 'Pasar' as the key field. The 'INNER' join type is selected in both.

- Kemudian pada Microsoft Excel Output setelah Multiway Merge Join, konfigurasi tetap dibiarkan default. Lakukan pada Microsoft Excel Output – Microsoft Excel Output 6

The image shows the 'Microsoft Excel output' configuration window. The 'File' tab is active, displaying fields for 'Filename' (set to 'file'), 'Extension' (set to 'xls'), and several checkboxes for including step name, date, and time in the filename. The 'Add filenames to result' checkbox is checked. The 'Date time format' field is empty. The 'Show filename(s)...' button is visible. The 'OK' and 'Cancel' buttons are at the bottom.



4. Select Values

Step selanjutnya ialah menghapus beberapa variabel *pasar* setelah dilakukan proses penggabungan. Variabel tersebut dihilangkan dan hanya disisakan pada bagian atas saja. Contoh konfigurasinya seperti pada gambar dibawah ini, lakukan pada Select Values – Select Values 6

The image shows three sequential screenshots of the 'Select values' dialog box, illustrating the process of selecting and renaming fields.

Screenshot 1: The 'Step name' is 'Select values'. The 'Fields' table lists fields 31 through 40. Field 33, 'Pasar_1', is selected. The 'Include unspecified fields, ordered by name' checkbox is unchecked.

#	Fieldname	Rename to	Length	Precision
31	30			
32	31			
33	Pasar_1			
34	1_1			
35	2_1			
36	3_1			
37	4_1			
38	5_1			
39	6_1			
40	7_1			

Screenshot 2: The 'Step name' is 'Select values'. The 'Fields' table lists fields 31 through 35. Field 32, '31', is selected. The 'Include unspecified fields, ordered by name' checkbox is unchecked.

#	Fieldname	Rename to	Length	Precision
31	30			
32	31			
33	1_1			
34	2_1			
35	3_1			

Screenshot 3: The 'Step name' is 'Select values'. The 'Fields' table lists fields 1 through 8. Field 1, 'Pasar', is selected. The 'Include unspecified fields, ordered by name' checkbox is unchecked.

#	Fieldname	Rename to	Length	Precision
1	Pasar			
2	1			
3	2			
4	3			
5	4			
6	5			
7	6			
8	7			



5. Merge Join

Step selanjutnya ialah melakukan penggabungan dari dua semester data yang sebelumnya sudah di bersihkan menggunakan Select Values. Konfigurasinya dapat dilihat pada gambar dibawah, lakukan pada Merge Join – Merge Join 3

The image shows three instances of the 'Merge join' dialog box. Each instance has the following fields:

- Step name: Merge Join, Merge Join 2, Merge Join 3
- First Step: Select values, Select values 3, Select values 5
- Second Step: Select values 2, Select values 4, Select values 6
- Join Type: INNER
- Keys for 1st step: # 1, Key field: Pasar
- Keys for 2nd step: # 1, Key field: Pasar
- Buttons: Get key fields, Help, OK, Cancel

6. Select Values

Lalu hapus variabel pasar dari hasil penggabungan dan sisakan variabel pasar pada list paling atas data seperti pada langkah 4 yang telah dilakukan sebelumnya. Lakukan pada Select Values 7 & Select Values 8

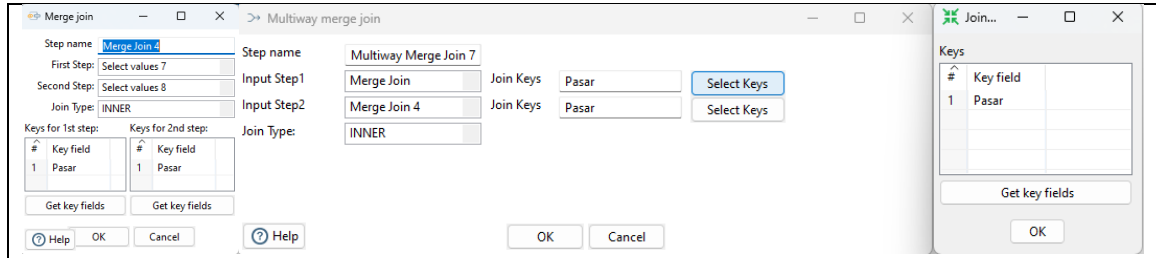
The image shows three instances of the 'Select values' dialog box. Each instance has the following fields:

- Step name: Select values 7, Select values 7, Select values 7
- Fields: # 1, Fieldname: 30_4, Rename to: Pasar_1, Length: 1, Precision: 6
- Buttons: Get fields to select, Edit Mapping, Help, OK, Cancel



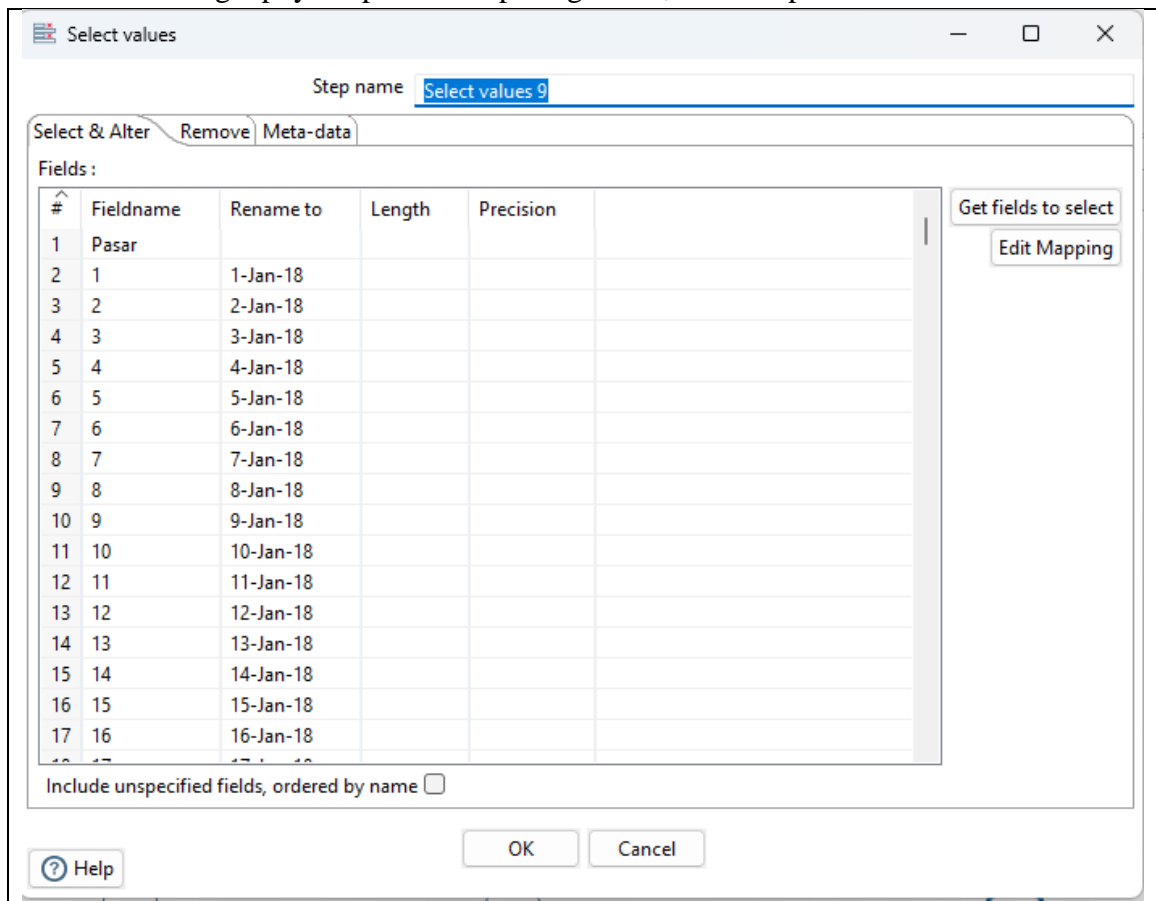
7. Merge Join & Multiway Merge Join

Step selanjutnya ialah melakukan Merge Join dan setelahnya melakukan Multiway Merge Join dengan konfigurasi sebagai berikut, lakukan pada Merge Join 4 & Multiway Merge Join 7



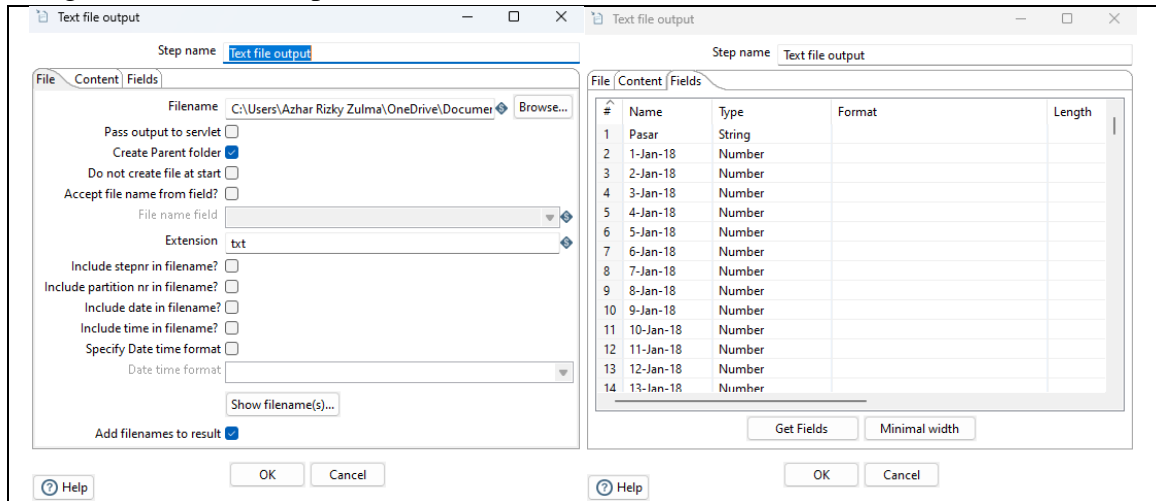
8. Select Values

Step selanjutnya ialah menghapus variabel pasar dari hasil penggabungan dan sisakan variabel pasar pada list paling atas data seperti pada langkah 4 dan langkah 6 yang telah dilakukan sebelumnya lalu rename setiap data dengan format *DD-MMM-YY* contoh seperti 19-Dec-20 selengkapnya dapat dilihat pada gambar, lakukan pada Select Values 9



9. Text File Output

Langkah terakhir simpan data menggunakan step *Text File Output* dengan konfigurasi sebagai berikut lalu simpan dengan format nama



2. Alat dan Bahan

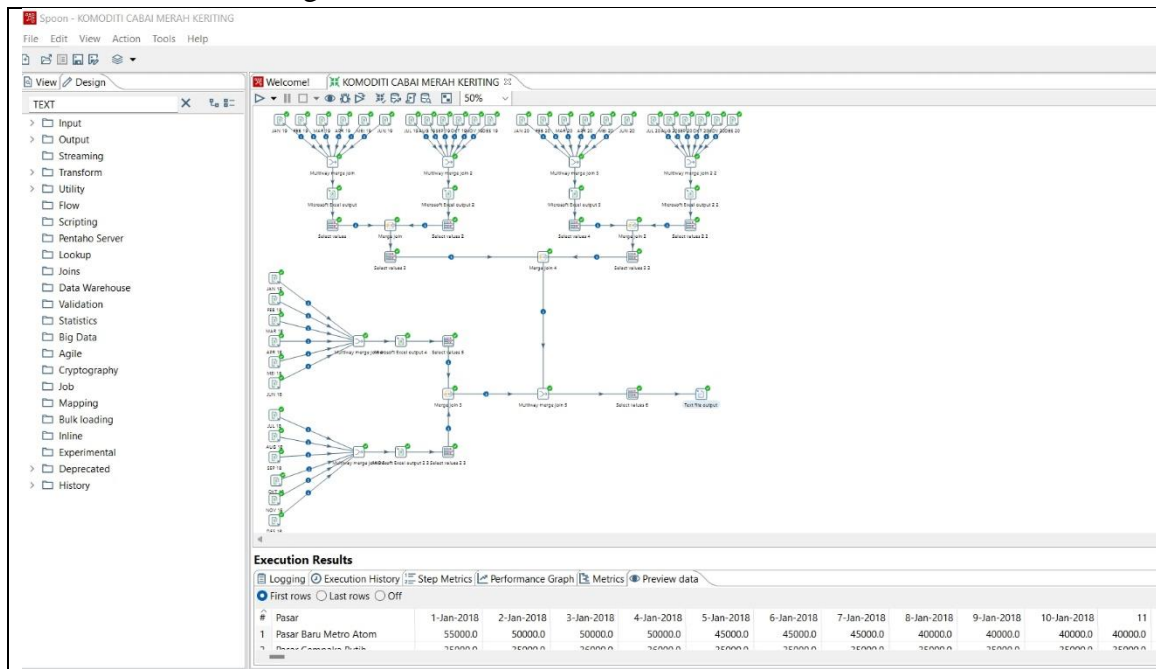
Hardware : Laptop/PC

Software : Spoon Pentaho from Hitachi Vantara

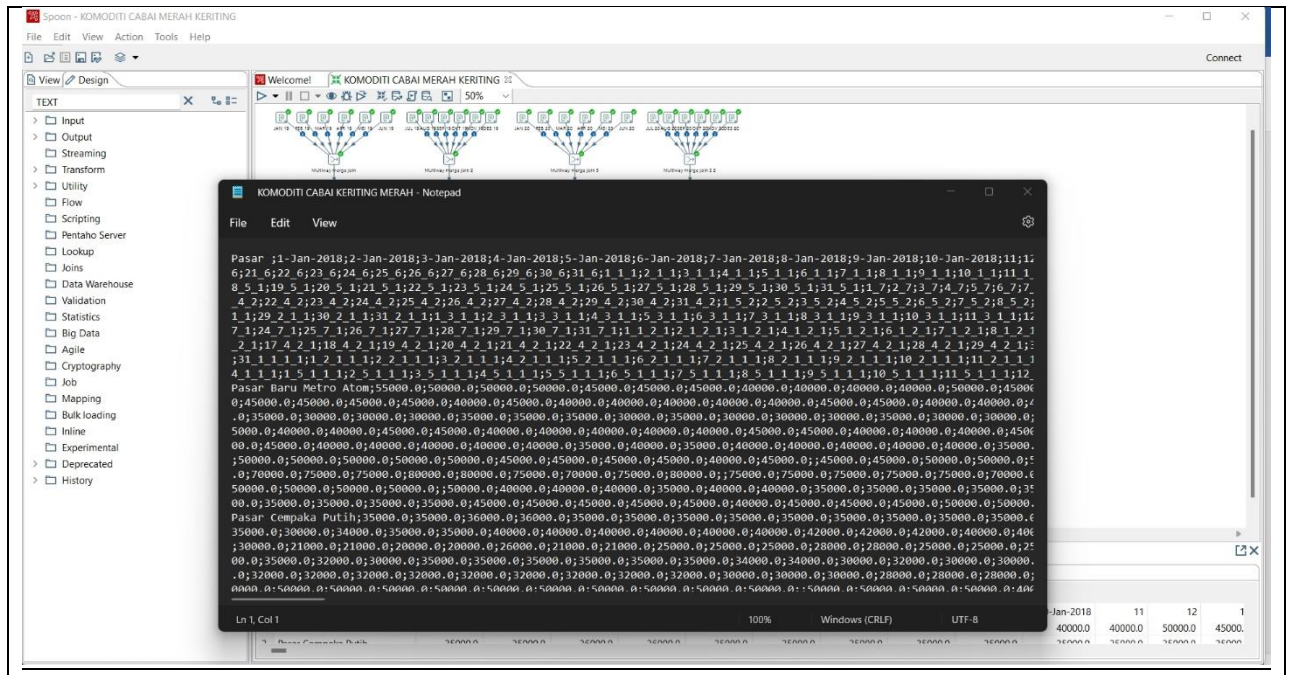
3. Elemen Kompetensi

a. Tugas Akhir Praktikum IV

1. Struktur ETL Gabungan



2. Data ETL Gabungan



4. File Praktikum

Github Repository:

<https://github.com/IbnuFajar7/Data-Warehouse/upload/main/Prak-12>

5. Soal Latihan

Soal:

1. Apa itu ETL?
2. Mengapa ETL itu penting?

Jawaban:

1. ETL adalah singkatan dari extract, transform, dan load. Merupakan proses integrasi data. Data akan dikombinasikan dari berbagai sumber.
2. Berperan dalam pengisian data untuk kebutuhan analisis dengan melakukan ekstraksi dari berbagai sumber data, pengubahan data ke bentuk yang sesuai dan pengisian ke storage data warehouse.



6. Kesimpulan

- Dalam pengerjaan praktikum Data Warehouse, kita harus benar-benar teliti dalam menginputkan suatu fungsi untuk menampilkan suatu keluaran pada layar dengan sesuai.
- Kita dapat mengetahui suatu kepentingan dari penggabungan data yang sudah kita cleansing, semuanya berujung kepada memudahkan kita untuk memberikan rangkuman data yang begitu banyaknya untuk lebih mudah dibaca dan dimengerti untuk khalayak ramai.

7. Cek List (✓)

No	Elemen Kompetensi	Penyelesaian	
		Selesai	Tidak Selesai
1.	Tugas Akhir Praktikum IV	✓	

8. Formulir Umpan Balik

No	Elemen Kompetensi	Waktu Pengerjaan	Kriteria
1.	Tugas Akhir Praktikum IV	30 Menit	1

Keterangan:

- Menarik
- Baik
- Cukup
- Kurang

